

BNSTs: In the eye of the beholder

Timothy J. Kaufmann[®] and Bradley J. Erickson

Department of Radiology, Mayo Clinic, Rochester, Minnesota, USA (T.J.K., B.J.E.)

Corresponding Author: Timothy J. Kaufmann, MD, MS, Department of Radiology, Mayo Clinic, 200 1st St. SW, Rochester, MN 55905, USA (kaufmann.timothy@mayo.edu).

As reported in the study by Dr Zhang et al,¹ it is the machine beholder who betters the human beholder in preoperatively classifying benign peripheral nerve sheath tumors (BNSTs) as neurofibroma or schwannoma, from a group of 107 schwannomas and 59 neurofibromas consecutively surgically resected at three academic institutions. Better by a lot? By enough. None of four experienced human observers alone, nor combined through one “expert score,” with or without clinical information, had accuracy significantly greater than the no-information rate of 0.643 (ie, the proportion of the largest class, namely schwannoma). But the machine did. For instance, the author’s logistic regression classifier, using both imaging and clinical features, delivered an accuracy of 0.929 and an AUC of 0.923. Not bad for a first attempt at this using a relatively small number of patients for a radiomics and machine-learning task.

Those who see BNST patients know the problem: a well-defined enhancing mass associated with a peripheral nerve is identified with cross-sectional imaging—with or without pain, with or without an ascribable neurologic deficit—but it’s notoriously difficult to differentiate neurofibroma from schwannoma by imaging. This matters because these two tumors have different growth patterns and therefore different surgical risks and odds of complete surgical resection.² Schwannomas are typically encapsulated and involve only a single nerve fascicle, but neurofibromas do not have a capsule and often involve many if not most of the fascicles within a nerve. Therefore, it is generally easier to surgically isolate and resect a schwannoma while sparing most of the nerve, but the resection of neurofibromas carries higher risk of permanent neurologic deficit. Unfortunately, percutaneous biopsy prior to resection also carries risk of neurologic deficit,² so this is not typically performed for presumed BNSTs.

There are distinctive imaging findings of the rare intraneuronal perineurioma,³ but qualitative imaging findings for discriminating between fibroma and schwannoma lack in sensitivity and specificity. It is difficult at imaging to resolve whether BNST involves just one fascicle or many. The “target sign” of central low signal intensity surrounded by a hyperintense rim on T2-weighted MRI has been described in neurofibroma,

related to a fibrocollagenous core surrounded by more myxomatous tissue, though this sign can also be seen with schwannoma if its core has more cellular Antoni A tissue surrounded by hypocellular Antoni B tissue.^{4,5} Of course, knowing the presence of an NF-1 or NF-2 or schwannomatosis diagnosis is very helpful, which would make one lean strongly toward neurofibroma in the case of NF-1 and toward schwannoma in the case of NF-2 or schwannomatosis. But particularly if a BNST is solitary and sporadic, it may be impossible to tell neurofibroma from schwannoma until the tumor is surgically exposed.

In the last decade, radiomics and machine learning and deep learning (these not being mutually exclusive in their use) have been used for many classification tasks (is it “A” or “B”?); the prediction of response to therapy, progression-free survival, and overall survival.^{6,7} Zhang et al trained and tested radiomics-based classifier models for neurofibroma vs schwannoma using machine learning. In this study, neuroradiologists segmented the tumors on T1-weighted post-gadolinium imaging, and the commonly used PyRadiomics library was then employed on these tumors to extract 900 quantitative imaging features (eg, shape and texture parameters) which are generally beyond the human eye’s capacity to discern. Knowing some clinical data, such as phakomatosis syndrome status and age, was strongly helpful in predicting tumor histology, but imaging data did add to the strength of their predictive models. As explained by the authors, it is encouraging that the radiomics signatures of spherical disproportion and kurtosis were consistently predictive of tumor type, given that these would fit with previously described radiographic patterns, suggesting that their models were not completely “black box.”

Some of the authors’ findings reflect the particularities of their patient groups. For instance, NF-1 was the strongest predictor in their models, but NF-2 and schwannomatosis status were dropped as predictors from models through feature reduction steps, likely due to the small number of NF-2 and schwannomatosis patients in their dataset. The small sample size may have limited the discovery of imaging features which would be important in the classification task. It may be that

imaging features would not significantly add to a machine learning classifier in syndromic patients where the clinical information of syndrome type dominates the prediction. But for spontaneous BPNST patients, imaging findings would be expected to dominate in those predictive models. There were not enough spontaneous BPNST patients in the authors' study to support a subgroup analysis comparing syndromic with non-syndromic BPNST tumors, though the authors do hope to later achieve this. At that point, we should know better the ability of imaging with machine learning to preoperatively classify a spontaneous BPNST as neurofibroma or schwannoma.

Artificial intelligence (AI), including machine learning, is notoriously hungry for data. Therefore, applying AI techniques to uncommon tumors will always be problematic. Throughout AI, there is the ubiquitous "large p, small n" challenge—a large number of radiomics features (p or number of columns in a data matrix) relative to a small number of imaging samples (n or number of rows). Here it is with BPNSTs as well. Other limitations include that the authors' models have not been tested on datasets outside of their three institutions nor with different proportions of syndromic vs non-syndromic tumors. We would not know if the resected tumors in their study might vary in any way detectable by machine learning from tumors that did not reach the clinical threshold which prompted resection. Problems with repeatability and generalizability of AI-based algorithms have been frequently encountered in radiology for various other reasons, including variability in imaging sequence acquisition and in preprocessing steps such as the normalization of image intensities. These and many other issues (not least being regulations and reimbursement) continue to hamper the translation of AI models into clinical practice.⁸

Nonetheless, Zhang et al solidly contribute to the ever-growing list of studies which show that machines can behold useful imaging information that the human eye cannot. The prediction of timelines for implementation into the clinic of AI tools such as this one seems fraught, but it

does seem inevitable that AI will eventually permeate our clinical practice as a very useful tool.

Acknowledgments

This text is the sole product of the authors and no third party had input or gave support to its writing.

References

1. Zhang M, Tong E, Wong S, et al. Machine learning approach to differentiation of peripheral schwannomas and neurofibromas: a multi-center study [published online ahead of print September 6, 2021]. *Neuro Oncol.* 2022;24(4):601–609.
2. Perez-Roman RJ, Shelby Burks S, Debs L, Cajigas I, Levi AD. The risk of peripheral nerve tumor biopsy in suspected benign etiologies. *Neurosurgery.* 2020;86(3):E326–E332.
3. Wilson TJ, Howe BM, Stewart SA, Spinner RJ, Amrami KK. Clinicoradiological features of intraneuronal perineuriomas obviate the need for tissue diagnosis. *J Neurosurg.* 2018;129(4):1034–1040.
4. Banks KP. The target sign: extremity. *Radiology.* 2005;234(3):899–900.
5. Bhargava R, Parham DM, Lasater OE, et al. MR imaging differentiation of benign and malignant peripheral nerve sheath tumors: use of the target sign. *Pediatr Radiol.* 1997;27(2):124–129.
6. Beig N, Bera K, Tiwari P. Introduction to radiomics and radiogenomics in neuro-oncology: implications and challenges. *Neurooncol Adv.* 2020;2(Suppl 4):iv3–iv14.
7. Bera K, Braman N, Gupta A, Velcheti V, Madabhushi A. Predicting cancer outcomes with radiomics and artificial intelligence in radiology. *Nat Rev Clin Oncol.* 2021. doi:10.1038/s41571-021-00560-7.
8. Lui YW, Chang PD, Zaharchuk G, et al. Artificial intelligence in neuroradiology: current status and future directions. *AJNR Am J Neuroradiol.* 2020;41(8):E52–E59.